Audiovisual correlates of basic emotions in blind and sighted people

Audiovisual correlates of basic emotions in blind and sighted people

Marc Swerts, Kitty Leuverink, Madelène Munnik and Vera Nijveld

Tilburg University, School of Humanities, TiCC research center, Tilburg (The Netherlands)

m.g.j.swerts@uvt.nl

Abstract

This study is concerned with the expression and recognition of basic emotions in blind and sighted people. We collected audiovisual recordings from blind and sighted people who were asked to produce specific utterances in such a way that they would fit different emotional contexts (i.e., angry, happy, sad, scared). In a perception experiment, 75 sighted participants had to guess which emotion from a blind or sighted person was enacted in one of three conditions (audio-only, video-only, audiovisual). While emotions expressed by sighted people were comparatively more easy to judge in audiovisual and video-only presentations, it turned out to be the case that emotions of blind people were more often correctly classified in the audio-only condition. Interestingly, the general patterns in classification accuracy were remarkably similar across conditions, and across speaker type.
Index Terms: blind and sighted people, audiovisual prosody, emotional expressions

1. Introduction

As is the case with many other forms of cognitive development (Goswami 2008), there has been quite some scholarly debate regarding the degree to which the acquisition of linguistics skills is a matter of nature or nurture: to what extent do people "naturally" learn a language merely from being exposed to linguistic input from their environment, and to what extent are they explicitly taught about linguistic structures by people around them. While this problem has been extensively addressed for a range of phenomena (lexical, phonetic, grammatical), similar questions have been asked about the acquisition of nonverbal features, where nonverbal features comprise both auditory (intonation, rhythm, loudness, ..) and visual forms (facial expressions, hand and arm gestures, ...).

One intriguing issue in particular is whether people learn to produce facial expressions from mimicking other people's behavior, or because they are genetically predisposed to produce such expressions, very much like they are borne with an ability to breath, swallow and suck. To answer this, people have been specifically interested in expressions produced by blind people (e.g. Dyck et al. 2002; Galati et al. 1997, 2001; Matsumoto et al. 2009; Conti-Ramsden & Pérez-Pereira 1999). Indeed, given that blind people have had no or only limited access to visual input during their lives, it is interesting to explore whether they produce any interpretable expressions at all. Previous research in this area has been focusing mainly on how such expressions might cue specific emotions (Ekman 1994), and revealed that blind people are indeed able to produce facial expressions for that particular purpose, suggesting that such expressive skills are innate.

In line with previous research, the current paper will address a number of questions that have not been properly answered in previous work. First, this study wants to shed light on the relation between auditory and visual cues to emotions in the expressions of blind and sighted people. The assumption here would be that blind people might compensate in their auditory cues what they have failed to learn to show through their facial expressions. The more specific question is whether their auditory cues would then become more easily interpretable for observers than the ones produced by sighted people. In addition, if it is indeed the case that blind people display their emotions through audiovisual expressions, it is interesting to find out whether their behavioral patterns are similar to those of sighted people, i.e., whether the success with which they display different emotions is comparable to how sighted people do this.

To investigate such questions, the current study explores dynamic stimuli (audiovisual recordings of utterances), which are arguably more ecologically valid than those of many previous studies (e.g. still images without sound). The current study consists of two parts. First, we use a specific elicitation procedure to record audiovisual expressions of basic emotions produced by blind and sighted people. For this, we let people act different versions of lexically identical utterances that fit different emotional contexts. Second, the utterances are presented to independent observers who have to rate the emotion expressed in the utterances, in one of three conditions (video-only, sound-only, video and sound).

2. Audiovisual recordings

2.1. Participants

Fifteen blind people (aged 34 to 74, M = 55.5, SD = 12.4) (7 men, 8 women) took part in the study. They were

Figure 1: *Stills of facial expressions of an angry, scared, sad and happy emotion produced by a blind (top) and sighted (bottom) person.*

all members of the NVBS, a nationwide organization for blind people in the Netherlands, and came from different regions of the Netherlands. Fourteen of the subjects were (functionally) blind since birth, the other one was blind since very early age. In addition, twenty sighted people (one male, nineteen females) took part in the experiment. The latter were all students at Tilburg University, who took part for course credit. More demographic information about them was missing, but they were all bachelor students in the teaching programme.

2.2. Procedure

Both blind and sighted participants had to produce utterances with a neutral content (e.g. *Mijn buurman is verhuisd* (My neighbour has moved)) in such a way that it would fit different contexts, i.e., sad, angry, scared and happy. To do this, the participants had to speak the sentence after an introduction in which the experimentor first mentioned a specific emotion (e.g. sad), and then described a specific context that matched that emotion. For instance, the "sad" context for the "neighbour" sentence was introduced by the following description: "He had always been like a friend to me. And now he is gone." The "happy" context for that sentence was introduced like this: "Finally I don't need to listen to that loud music anymore! I finally am able to sleep well!". Along the same lines, we created contexts for all 4 emotions for 4 different sentences, leading to 16 recordings of target utterances per speaker. Care was taken to present the spoken context in a neutral manner to the participant in order not induce a specific style in the participant. The reason to present participants with spoken contexts (rather than written ones) was that this more easy to implement for our blind participants. The complete procedure was recorded on videotape with permission of the subjects. Figure 1 shows some stills of a blind and sighted person taken from utterances in the various emotional contexts.

2.3. Annotations

To get a first impression of the kinds of features blind and sighted people display when producing the various emotional utterances, their facial characteristics were annotated. Two independent coders broadly transcribed the facial properties of the clips in terms of presence or absence of a number of clear features (e.g. speaker frowns, speaker looks down, speaker moves lip corners up, ...). In cases where the kappa-score between the two coders was not sufficiently high, the presence or absence of a feature was decided by consensus labeling. The annotations revealed that blind and sighted people are remarkably similar in the kinds of facial features they use in the various emotional contexts. To give a few conspicuous examples (see also Figure 1): For both kinds of speakers, it turned out to be the case that the happy emotion was associated more often with a prototypical smile (lip corners up) and lifted eyebrows, whereas expressions with the lip corners down occurred far less in the happy emotion than in the others. Sad and angry emotions were more often displayed with eyebrow frowns than the other emotions. Both kinds of speakers more often turned their heads down when displaying a sad emotion. While these annotations only gave a first indication regarding possible

Table 1: Mean proportions of correct classifications for emotions for sighted and blind people. Standard errors are written between brackets.

Condition	Emotion	Sighted	Blind
Video-only	Scared	.36 (.13)	.26 (.10)
	Angry	.45 (.10)	.22 (.11)
	Sad	.61 (.11)	.53 (.11)
	Happy	.91 (.06)	.62 (.15)
Audio-only	Scared	.31 (.09)	.39 (.13)
	Angry	.39 (.11)	.59 (.15)
	Sad	.75 (.14)	.63 (.12)
	Happy	.85 (.09)	.84 (.10)
Audiovisual	Scared	.49 (.18)	.35 (.12)
	Angry	.57 (.09)	.44 (.10)
	Sad	.76 (.11)	.67 (.11)
	Happy	.95 (.05)	.81 (.15)

visual cues, they do suggest that blind people indeed use similar facial expressions as sighted people to distinguish a few basic emotions. Given this, we set up a perception experiment to investigate the cue value of such expressions, and how they relate to the auditory features that speakers also use to transmit specific emotions.

3. Perception test

3.1. Participants

Seventy-five subjects took part in the experiment. Those were all students at Tilburg University. From two subjects, the demographic information was missing. The mean age of the other seventy-three subjects was 21.9 (SD = 4.4), ranging from 18 to 48. Twenty-one (28.8%) were male, fifty-two (71.2%) were female. They were randomly assigned to one of three conditions (audio-only, video-only, audiovisual).

3.2. Procedure

Judges were shown video clips of the sentence "Mijn buurman is verhuisd" produced in four emotional contexts (sad, angry, happy, scared). First, judges were shown the set of utterances expressed by sighted people, followed by the set of utterances expressed by blind people. The stimuli of the blind speakers were shown after those of the sighted people, as these were hypothesized to be more difficult to judge. Two sets, one containing sixty utterances of the target sentence expressed by sighted people, and one containing all forty utterances of this sentence expressed by blind people were used. Within these two sets, utterances were randomized. Participants were asked to judge which emotion was being expressed. They could fill in their answer on a form as a multiple-choice task where they had to choose between 4 emotions. None of the participants had acted as speakers in the recording sessions.

Table 2: Mean proportions of correct classifications of emotions for sighted and blind people. Standard errors are written between brackets.

Emotion	Sighted	Blind
Scared	.39 (.15)	.33 (.13)
Angry	.47 (.12)	.42 (.19)
Sad	.70 (.14)	.61 (.13)
Happy	.90 (.08)	.76 (.17)

3.3. Results

We first analysed the responses by conducting a repeated measures ANOVA with condition (audio-only, video-only, audiovisual) as between-subject factor, with sight (blind, sighted) and emotion (sad, scared, happy, angry) as within-subject factors, and with the proportion of correctly guessed emotions as dependent variable. Table 1 gives the proportion of correct responses for the factors condition, emotion and sight. Note that almost all cells have numbers above chance level (.25). There was a main effect of sight: $F_{(1,72)} = 118.229, p < .001, \eta_p^2 = .622$. Judges tended to give more correct answers for stimuli produced by sighted people (M = .61, SE = .10) than for those by blind people (M = .53, SE = .10). There was also a main effect of emotion: $F_{(3,216)} = 396.788, p < .001, \eta_p^2 = .846$. Happiness was most often guessed correctly (M = .83, SE = .10), followed by sadness (M = .66, SE = .10), anger (M = .44, SE = .10) and scared (M = .36, SE = .10). Finally, we found a main effect for condition: $F_{(2,72)} = 60.302, p < .001, \eta_p^2 = .636$. The clips in the audiovisual condition got the most correct answers (M = .63, SE = .10), followed by the clips in the audio-only condition (M = .59, SE = .10) and the video-only conditions (M = .49, SE = .10).

In addition, we found a significant 2-way interaction effect between sight and emotion, even when the effect size is relatively small: $F_{(3,216)} = 7.031, p < .001, \eta_p^2 = .089$. As table 2 shows, the response patterns for the different emotions for data from sighted and blind people are very similar, but the difference in scores between the emotions is larger for the stimuli from sighted people than from blind people. More interestingly, there was also a significant interaction between sight and condition: $F_{(2,72)} = 64.729, p < .001, \eta_p^2 = .643$ (see also Table 3). In the video-only and audiovisual conditions, people tended to give more correct answers for data of sighted people (M = .58, SE = .10 and M = .69, SE = .10, respectively) than for those of blind people (M = .40, SE = .13 and M = .57, SE = .13 respectively). In the audio-only condition, however, judges were better in judging the emotions of blind people (M = .61, SE = .13) than those of sighted people (M = .57, SE = .10).

Finally, it is revealing to look at the way emotions were confused in the perception task. Overall, it turned out that the three negative emotions got poorer scores

Table 3: Mean proportions of correct classifications for emotions in different conditions for sighted and blind people. Standard errors are written between brackets.

Condition	Sighted	Blind
Visual	.58 (.10)	.40 (.13)
Auditory	.57 (.10)	.61 (.13)
Audiovisual	.69 (.10)	.57 (.13)

than the positive one, both for data from blind and sighted people. For the sighted actors, fear and anger were confused with sadness in all three presentation conditions. For the blind actors, fear and anger were confused with sadness in the video-only condition, whereas in the audio-only and audiovisual presentation conditions, fear was only confused with sadness, for data of blind people.

4. General discussion

This study was concerned with the expression and recognition of basic emotions in blind and sighted people. We collected audiovisual recordings from blind and sighted people who were asked to produce specific utterances in such a way that they would fit different emotional contexts (i.e., angry, happy, sad, scared). In a perception experiment, 75 sighted participants had to guess which emotion from a blind or sighted person was enacted in one of three conditions (audio-only, video-only, audiovisual). While emotions expressed by sighted people were comparatively more easy to judge in audiovisual and video-only presentations, it turned out to be the case that emotions of blind people were more often correctly classified in the audio-only condition.

Interestingly, the general patterns in classification accuracy were remarkably similar across conditions, and across speaker type. Both for blind and sighted people it was the case that the happy emotion is most clearly expressed, and the emotions sad and scared are confused more often with each other than with other emotions. It is hard to speculate on why the distributional results turned out to be like this. It could be that happiness has clearer audiovisual correlates simply because people are more experienced in showing this emotion, being more socially acceptable than showing the other ones. However, that would be in conflict with the idea that showing emotions like anger would be more important from an evolutionary perspective (Ekman 1994). Maybe these results are simply due to the fact that happiness is more easily interpretable because it was the only positive emotion, and in that sense more clearly distinguishable from the other three that are all negative. In any case, this then begs the question as to how blind and sighted people signal and interpret a range of other emotions. Those could also include what has been termed social emotions, i.e., emotions that regulate social interactions between people, and that often tend to differ between cultures and settings. In that respect, we are setting up research in which we aim to investigate how people show their confidence level about an answer they give to an easy or difficult question. As we have argued before (e.g. Swerts & Krahmer 2005), the expression of (un)certainty is a typically social skill that people need to acquire as part of their development, as they grow older. To investigate this, one would have to exploit methods that are more ecologically valid than the acting procedure we used here (Swerts & Krahmer 2008; Krahmer & Swerts 2011). Related to this, note that the expressions people displayed in our experiment were not spontaneous, but posed (acted). In the future, it would be interesting to explore how results from the current investigation generalize to more natural settings.

5. Acknowledgments

We would like to thank the "Nederlandse Vereniging van Blinden en Slechtzienden" (NVBS), a Dutch, nationwide organization for blind people, for their support in this project, and for their help in finding blind participants for our experiments.

6. References

Conti-Ramsden, G., & Pérez-Pereira, M. (1999).*Language development and social interaction in blind children.* Hove: Psychology Press.

Dyck, M., Farrugia, C., Shochet, I., & Holmes-Brown, M. (2002). Emotion recognition ability in deaf or blind children: Do sounds, sights, or words make the difference? *Journal of Child Psychology and Psychiatry*, 45(4).

Ekman, P. (1994). *The nature of Emotion*. New York: Oxford University Press

Galati, D., K.R. Scherer & P.E. Ricci-Bitti (1997). Voluntary Facial Expression of Emotion: Comparing Congenitally Blind With Normally Sighted Encoders. *Journal of Personality and Social Psychology* 73(6), 1363-1379.

Galati, D., Miceli, R. & Sini, B. (2001). Judging and coding facial expression of emotion in congenitally blind children. *International Journal of Behavioral Development*, 25(3), 268-278.

Goswami, U. (2008).*Cognitive development. The learning brain.* Hove and New York: Psychology Press.

Krahmer, E. & Swerts, M. (2011). Audiovisual expression of emotions in communication. In: *Probing Experience II* (ed. by M. Ouwerkerk, M. Krans & J. Westerink), Philips Research Book Series, Springer.

Matsumoto, D., & Willingham, B. (2009). Spontaneous Facial Expressions of Emotion of Congenitally and Noncongenitally Blind Individuals. *Journal of Personality and Social Psychology*, 96(1).

Swerts, M. & Krahmer, E. (2008) Gender-related differences in the production and perception of emotion. *Proc. Interspeech 2008*, Brisbane, Australia, September 2008.

Swerts, M. & Krahmer, E. (2005) Audiovisual prosody and feeling of knowing. *Journal of Memory and Language* 53, pp. 81-94.